

# METHOD AND SYSTEMS FOR PREDICTION OF HLA CLASS II-SPECIFIC EPITOPES AND CHARACTERIZATION OF CD4+ T CELLS

## CROSS-REFERENCE

**[0001]** This application is a continuation of International Application No. PCT/US2019/068084 filed Dec. 20, 2019 which claims the benefit of U.S. Provisional Application No. 62/891,101, filed on Aug. 23, 2019; U.S. Provisional Application No. 62/855,379, filed on May 31, 2019; U.S. Provisional Application No. 62/826,827, filed on Mar. 29, 2019; and 62/783,914, filed on Dec. 21, 2018; each of which is incorporated herein by reference in its entirety.

## SEQUENCE LISTING

**[0002]** The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Jan. 31, 2020, is named 50401-735\_301\_SL.txt and is 27,415 bytes in size.

## BACKGROUND

**[0003]** The major histocompatibility complex (MHC) is a gene complex encoding human leukocyte antigen (HLA) genes. HLA genes are expressed as protein heterodimers that are displayed on the surface of human cells to circulating T cells. HLA genes are highly polymorphic, allowing them to fine-tune the adaptive immune system. Adaptive immune responses rely, in part, on the ability of T cells to identify and eliminate cells that display disease-associated peptide antigens bound to human leukocyte antigen (HLA) heterodimers.

**[0004]** In humans, endogenous and exogenous proteins can be processed into peptides by the proteasome and by cytosolic and endosomal/lysosomal proteases and peptidases and presented by two classes of cell surface proteins encoded by MHC genes. These cell surface proteins are referred to as human leukocyte antigens (HLA class I and class II), and the group of peptides that bind them and elicit immune responses are termed HLA epitopes. HLA epitopes are a key component that enables the immune system to detect danger signals, such as pathogen infection and transformation of self. CD4+ T cells recognize class II MHC (HLA-DR, HLA-DQ, and HLA-DP) epitopes displayed on antigen presenting cells (APCs), such as dendritic cells and macrophages. The endogenous processing and presentation of HLA class II-ligands is a complex procedure and involves a variety of chaperones and a subset of enzymes that are not all well characterized. HLA class II-peptide presentation activates helper T cells, subsequently promoting B cell differentiation and antibody production as well as CTL responses. Activated helper T cells also secrete cytokines and chemokines that activate and induce differentiation of other T cells.

**[0005]** Understanding the peptide-binding preferences of every HLA class II heterodimer is the key to successfully predicting which cancer or tumor-specific antigens are likely to elicit the cancer or tumor-specific T cell responses. There is a need for methods of identifying and isolating specific HLA class II-associated peptides (e.g., neoantigen peptides). Such methodology and isolated molecules are useful, e.g., for the development of therapeutics, including but not limited to, immune based therapeutics.

## SUMMARY

**[0006]** The methods and compositions described herein find uses in a wide range of applications. For example, the methods and compositions described herein can be used to identify immunogenic antigen peptides and can be used to develop drugs, such as personalized medicine drugs, and isolation and characterization of antigen-specific T cells.

**[0007]** CD4+ T cell responses may have anti-tumor activity. A high rate of CD4+ T cell responses may be shown without using Class II prediction (e.g., 60% of SLP epitopes in NeoVax study (49% in NT-001, see Ott et al., Nature, 2017 Jul. 13; 547(7662):217-221), and 48% of mRNA epitopes in Biontech study, see Sahin et al., Nature, 2017 Jul. 13; 547(7662):222-226). It may not be clear whether these epitopes are typically presented natively (by tumor or by phagocytic DCs). It may be desirable to translate high CD4+ T response rates into therapeutic efficacy by improving identification of truly presented HLA class II binding epitopes.

**[0008]** The roles of gene expression, enzymatic cleavage, and pathway/localization bias may have not been robustly quantified. It may be unclear whether autophagy (HLA class II presentation by tumor cells) or phagocytosis (HLA class II presentation of tumor epitopes by APCs) is the more relevant pathway, although most existing MS data may be presumed to derive from autophagy. NetMHCIIpan may be the current prediction standard, but it may not be regarded as accurate. Of the three HLA class II loci (DR, DP, and DQ), data may only exist for certain common alleles of HLA-DR.

**[0009]** There may be different data generation approaches for learning the rules of HLA Class II presentation, including the field standard and the proposed approach. The field standard may comprise affinity measurements, which may be the basis for the NetMHCIIpan predictor, providing low throughput and requiring radioactive reagents, and it misses the role of processing. The proposed approach may comprise mass spectrometry, where data from cell lines/tissues/tumors may help determine processing rules for autophagy and mono-allelic MS may enable determination of allele-specific binding rules (multi-allelic MS data is presumed overly complex for efficient learning (Bassani-Sternberg. MCP. 2018)).

**[0010]** There may be different ways to validate the new HLA class II predictors: validation on held-out MS data, which may be default setting; retrospective of vaccine studies (e.g. NT-001), where immune monitoring data may assess vaccine peptide loading on APCs rather than tumor presentation and data may be thinly stretched across many different alleles; biochemical affinity measurements, which may be configured to get measurements for discordantly predicted peptides (only for 2-3 alleles); T cell inductions, which may be configured to test the rates at which Neon-preferred and NetMHCIIpan-preferred epitopes induce ex vivo T cell responses.

**[0011]** For validation through T cell inductions, the default approach may comprise assessing neoORFs from TCGA that are discordantly predicted, wherein induction materials may comprise healthy donor APCs and T cells and induction and readout may be via SLP (~15mer peptides). Random peptides may give a high rate of responses and SLP may insufficiently address processing. Possible solutions may comprise induction via mRNA.